

LAPORAN
TUGAS AKHIR



Disusun Oleh :

FITRALOKA ARUM SARI

0734010186

JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2012

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmatnya kepada penulis sehingga penulis dapat menyelesaikan Tugas Akhir dengan baik dan benar.

Penyusun menyadari bahwasanya dalam menyelesaikan Skripsi ini telah mendapat banyak bantuan dan dukungan dari berbagai pihak, untuk itu pada kesempatan yang berharga ini, penyusun mengucapkan terimakasih kepada :

1. Bapak Ir. Sutiyono, MT selaku Dekan FTI, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Ni Ketut Sari, MT selaku Ketua Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Sri Redjeki, MT selaku Dosen Pembimbing I Skripsi atas segala do’a restu, arahan, ilmu, dan bimbingan-bimbingan yang telah diberikan selama penyelesaian Skripsi.
4. Bapak S. Kom selaku Dosen Pembimbing II Skripsi atas segala do’a restu, arahan, ilmu, dan bimbingan-bimbingan yang telah diberikan selama penyelesaian Skripsi.
5. Bapak Basuki Rahmat, S. Si, MT, Ibu Rr. Ani Diah Raharjo, ST, M. Cs, Bapak Rizky Parlika, S. Kom selaku Dosen Penguji Seminar TA yang banyak memberi masukan yang berguna kepada penulis selama revisi.
6. Bapak dan Ibu dosen Teknik Informatika yang telah memberikan ilmunya kepada penulis selama kuliah.

7. Kedua orang tua tercinta yang selalu memberi dukungan dan doa supaya skripsi ini segera terselesaikan.
8. Keluarga Adi Hermawan, yang selalu membantu pengerjaan aplikasi ini dan memberikan masukan guna untuk memperlancar skripsi
9. Kakak-kakakku mas Gambar, mas Erik dan adik tersayang Trias, terima kasih atas do'a serta dukungannya selama proses mengerjakan Skripsi ini dan aku sayang kalian semua.
10. Keluarga Luluk Suharwati, yang mau meminjamkan print untuk penyusunan cetak skripsi ini dan doa yang selalu terpanjatkan.
11. Teman-teman Teknik Informatika angkatan 2011 rhina, lian, nanda, novita, hayu, chanif, faiq, didit, aldo, novihendra, sanggra, rizal, atik, widya, adidll. Teman bermain Randie, fika, tya, upid, andreas, novi, irul, erlinda dan semua orang yang berhubungan baik dengan penyusun yang tidak bisa disebutkan satu persatu terima kasih atas do'ad dan dukungannya selama ini serta terima kasih atas pertemanannya.
12. Enrique Iglesias dan Bruno Mars yang lagunya selalu menemani penyusunan dan pembuatan skripsi ini.
13. Yang paling berjasa tapi tidak merasa paling berjasa My Laptop Asus yang sering penyusun pakai di kala aplikasi mengalami error. Super Terima kasih.

MUCHAS GRACIAS.....!

Judul : Membuat aplikasi temu kembali informasi dengan menggunakan model ruang vektor pada
Pembimbing I : Prof.Dr.Ir.Sri Redjeki,MT
Pembimbing II : Faisal Muttaqin,S.Kom
Penyusun : Fitraloka Arum Sari

ABSTRAK

Pengguna Komputer dan berbagai aplikasinya banyak menyebabkan terjadinya penumpukan data dalam jumlah cepat yang sering kali menimbulkan kendala dalam mencari data yang dibutuhkan secara cepat, tepat dan tidak dapat dilakukan karena harus mencari secara manual atau satu per satu sehingga membutuhkan waktu cukup lama.

Kendala tersebut dapat bertambah apabila data yang dicari terletak di dalam jaringan komputer lokal sehingga lokasi data semakin tersebar, yang menyebabkan bertambahnya dan tenaga.

Untuk mengefisienkan hal tersebut diperlukan suatu aplikasi yang dapat melakukan pencarian secara cepat dan akurat dalam menemukan kembali data yang dibutuhkan, yang mampu mencari data di komputer stand alone maupun dalam jaringan komputer lokal.

Kata kunci : Temu kembali Informasi, Model Ruang Vektor

KATA PENGANTAR

Alhamdulillah, Penulis bersyukur kepada Allah SWT atas semua Rahmat, Berkah, dan Ridho-Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan Skripsi yang berjudul “Membuat aplikasi temu kembali informasi menggunakan model ruang vektor” ini dengan baik.

Skripsi merupakan salah satu syarat bagi mahasiswa untuk menyelesaikan program studi Sarjana Strata Satu (S1) di Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Melalui Skripsi ini penyusun merasa mendapatkan kesempatan besar untuk memperdalam ilmu pengetahuan yang diperoleh selama di bangku perkuliahan. Namun, penyusun menyadari bahwa Skripsi ini masih jauh dari sempurna. Oleh karena itu penyusun sangat mengharapkan saran dan kritik dari para pembaca untuk pengembangan aplikasi lebih lanjut.

Surabaya, 5 Juni 2012

(Fitraloka arum sari)

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	v
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan	5
1.5 Manfaat	6
1.6 Metodologi Penelitian	6
1.7 Sistematika penulisan	6
BAB II TINJAUAN PUSTAKA	9
2.1 Konsep Dasar Sistem Informasi	9
2.1.1 Pengertian Sistem Kembali Informasi	9
2.2 Pembobotan Kata	15
2.2.1 Pemodelan Ruang Vektor	16
2.3 Text Mining	18
2.4 Pengembangan Perangkat Lunak	19
2.5 Pengujian Perangkat Lunak	20
2.6 Data Flow Diagram	21
2.7 Definisi Pemrograman Delphi	21
2.7.1 Kelebihan Pemrograman Delphi	22
2.7.2 Membuat Sebuah Form di Delphi	24
2.7.3 Mengganti Nama Form dan Menambahkan Judul	25
2.8 Sejarah Borland Delphi	26
2.8.1 Sequence Diagram	27
2.8.2 Collaboration Diagram	27

2.8.3 Calss Diagram	27
2.9 My SQL	27
BAB III ANALISIS DAN PERANCANGAN SISTEM ...	30
3.1 Analisis Sistem	30
3.2 Analisis Kebutuhan	30
3.2.1 Fungsi Perangkat Lunak	30
3.2.2 Batasan perangkat lunak	31
3.2.3 Batasan perangkat Lunak	31
3.3 Model Fungsional perangkat Lunak	32
3.3.1 Data Flow Diagram	32
3.4 perancangan Perangkat Lunak	34
3.4.1 perancangan Arsitektur	34
3.4.2 Perancangan Data	36
3.4.3 Perancangan Antarmuka Perangkat Lunak	36
3.4.4 Perancangan Prosedural	37
3.4.5 Algoritma dan Flowchart	38
BAB IV IMPLEMENTASI DAN PENGUJIAN	44
4.1 Lingkungan pemrograman	44
4.1.1 Lingkungan perangkat keras	44
4.1.2 Lingkungan perangkat Lunak	44
4.2 Implementasi Arsitektur	45
4.3 Implementasi Proses	45
4.3.1 Implementasi Koneksi Database	46
4.4 Implementasi Aplikasi Desain Antar Muka	46
4.4.1 Form Halaman Database	51
4.5 Pengujian	52
4.5.1. Tujuan pengujian	53
4.5.2 Pengujian Perankingan	53
4.5.3 Kriteria pengujian	53
4.5.4 hasil pengujian	53
BAB V UJI COBA DAN EVALUASI PROGRAM	55
5.1 Uji coba Sistem	55

5.2 Uji coba Halaman Utama	56
5.2.1 Uji coba Tampilan Pencarian Lokasi File	56
5.2.2 Uji coba Tampilan Pencarian kata/ keyword	59
5.2.3 Uji coba Tampilan Perankingan	62
5.3 Form Halaman Database	64
BAB VI PENUTUP	64
6.1.Kesimpulan	66
6.2. Saran	66
DAFTAR PUSTAKA	68

DAFTAR GAMBAR

Gambar 2.1 Komponen Sistem Temu-Kembali Informasi (Tarto, 2008).....	13
Gambar 2.2 Interaksi antara pengguna dengan sistem (Tarto, 2008).....	13
Gambar 2.3 proses filter dokumen 2 tingkat.....	17
Gambar 2.4 Filter dokumen dengan model ruang vektor	18
Gambar 3.1 Diagram Konteks	32
Gambar 3.2 Diagram Aliran Data Level 1	33
Gambar 3.3 Rancangan antarmuka Utama perangkat Lunak	36
Gambar 3.4 Struktur Modul Perangkat Lunak	38
Gambar 3.5 Flowchart Perangkat Lunak	39
Gambar 3.6 Flowchart proses pencarian lokasi file	40
Gambar 3.7 flowchart proses pencarian nilai bobot dokumen	41
Gambar 3.8 flowchart perankingan	42
Gambar 4.1 Script Koneksi Database	46
Gambar 4.2 Tampilan Awal Aplikasi	47
Gambar 4.3 halaman hasil pencarian	48
Gambar 4.4 halaman pencarian lokasi file	49
Gambar 4.5 tampilan hasil daftar file dokumen.....	49
Gambar 4.6 halaman hasil perhitungan bobot dan perankingan	50
Gambar 4.7 Form Halaman Setting Database	51
Gambar 4.8 Form tabel Database tfile	52
Gambar 4.9 Form tabel Database tnilai	52
Gambar 4.10 Hasil Pengujian	54
Gambar 5.1 Uji Coba Tampilan Aplikasi	56
Gambar 5.2 Uji Coba Tampilan Pencarian lokasi file	57
Gambar 5.3 folder	58

Gambar 5.4 Uji Coba Tampilan Daftar file dokumen	58
Gambar 5.5 Uji Coba Tampilan Pencarian Kata/Keyword	59
Gambar 5.6 Uji Coba Tampilan Hasil Pencarian Kata/Keyword	60
Gambar 5.6 Uji Coba Tampilan Hasil isi dari file	61
Gambar 5.7 Uji Coba Tampilan Hasil isi dari perankingan	62
Gambar 5.8 Perankingan	63
Gambar 5.9 Nilai Bobot	63
Gambar 5.10 Form Halaman Setting Database	64
Gambar 5.11 Form tabel Database tfile	65
Gambar 5.12 Form tabel Database tnilai	65

DAFTAR TABEL

Tabel 2.1 Contoh frekuensi kata dalam suatu dokumen	16
Tabel 3.1 Keterangan Antar Muka Perangkat Lunak	37

BAB I

P E N D A H U L U A N

1.1. Latar Belakang

Semakin canggihnya teknologi di bidang komputasi dan telekomunikasi pada masa kini, membuat informasi dapat dengan mudah didapatkan oleh banyak orang. Kemudahan ini menyebabkan informasi menjadi semakin banyak dan beragam. Informasi dapat berupa dokumen, berita, surat, cerita, laporan penelitian, data keuangan, dan lain-lain. Tidak dapat dipungkiri lagi informasi telah menjadi komoditi yang paling penting dalam dunia modern masa kini.

Seiring dengan perkembangan informasi, banyak pihak menyadari bahwa masalah utama telah bergeser dari cara mengakses informasi menjadi memilih informasi utama yang berguna secara selektif. Usaha untuk memilih informasi ternyata lebih besar dari sekedar mendapatkan akses terhadap informasi. Pemilihan atau penemuan kembali informasi ini tidak mungkin dilakukan secara manual karena kumpulan informasi yang sangat besar dan terus bertambah besar.

Mandala (2002, hal 1) menyatakan bahwa suatu sistem otomatis diperlukan informasi (information retrieval system) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari pengguna secara otomatis berdasarkan kesesuaian dengan query (masukan berupa ekspresi kebutuhan informasi oleh pengguna) dari suatu koleksi informasi.

Namun dari hasil temuan kembali dokumen, pengguna tidak dapat melihat peran dari query dalam proses pencarian dokumen tersebut, urutan dokumen yang di-retrieve dinilai tidak informatif berdasarkan tingkat kesesuaiannya dengan query. Oleh karena itu digunakan model ruang vektor (vector space model) sistem temu kembali informasi. Menurut Arifin (dalam Salton, 1989) 'Salah satu model sistem temu kembali informasi yang paling sederhana namun paling produktif adalah model ruang vektor. Vektor model ini merepresentasikan term yang terdapat pada dokumen dan query. Elemen vektor tersebut adalah bobot term yang menjadi dasar penilaian dalam pemeringkatan dokumen. Hal yang perlu diperhatikan dalam penemuan kembali informasi model ruang vektor ini adalah pembobotan term (term weighting). Term dapat berupa kata, frase, atau unit hasil indexing lain dalam suatu dokumen sebagai gambaran konteks dari dokumen tersebut. Karena tiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka diperlukan indikator yaitu term weight (bobot term) dalam proses pencocokan dan perankingan dokumen terhadap query.

Metode pembobotan yang umumnya diunggulkan dalam penelitian-penelitian untuk digunakan dalam model ruang vektor yaitu TF-IDF Term Frequency Inverse Document Frequency (Arifin, 2002). Menurut Arifin (dalam Salton, 1989) 'Dalam perhitungan bobot term, sekalipun term frequency banyak digunakan, namun ia hanya mendukung proporsi jumlah dokumen yang dapat ditemukan-kembali oleh proses pencarian pada sistem Information Retrieval, sedangkan proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan pengguna akan lebih meningkat bila vektor bobot tersebut menggunakan term yang jarang muncul pada koleksi dokumen. Tentunya term

demikian akan diharapkan mampu mengelompokkan sejumlah dokumen yang memuatnya, sehingga berbeda dengan seluruh anggota koleksi dokumen lain yang tidak memilikinya. Kriteria ini dapat diakomodasi dengan menghitung invers frekuensi dokumen. Dengan digabungkannya kedua metode ini yaitu konsep frekuensi kemunculan term dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut, akan mampu meningkatkan proporsi jumlah dokumen yang dapat ditemukan kembali dan yang dianggap relevan secara sekaligus. Sehingga kriteria term yang paling tepat adalah term yang sering muncul dalam dokumen secara individu, namun jarang dijumpai pada dokumen lainnya.' Menurut Defeng (dalam Robertson, 2004: 503) 'Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen.

Selama ini dokumen (information retrieval) pada komputer pencari adalah berdasarkan kata kunci (keyword) sebagai input untuk proses pencarian. Ada banyak metode yang digunakan untuk proses pencarian ini, antara lain: term weighting, boolean model, dan vector model. Hasil dari pencarian dokumen

dengan menggunakan metode ini kadang belum tentu sesuai yang diinginkan, meskipun dalam dokumen yang diperoleh tersebut sudah mengandung kata kunci dalam masukan keyword. Salah satu penyebabnya adalah adanya berbagai konsep domain pengetahuan yang berbeda dalam kata kunci yang sama (problem semantik).

Penggunaan komputer dan berbagai aplikasinya banyak menyebabkan terjadinya penumpukan data dalam jumlah cepat yang sering kali menimbulkan kendala dalam mencari data yang dibutuhkan secara cepat, tepat dan tidak dapat dilakukan karena harus mencari secara manual atau satu per satu sehingga dapat membutuhkan waktu yang cukup lama.

Kendala tersebut dapat bertambah apabila data yang dicari terletak di dalam jaringan komputer lokal sehingga lokasi data semakin tersebar, yang menyebabkan bertambahnya waktu dan tenaga untuk mengefesienkan hal tersebut diperlukan suatu aplikasi yang dapat melakukan pencarian secara cepat dan akurat dalam menemukan kembali data yang dibutuhkan, yang mampu mencari data di komputer stand alone maupun dalam jaringan komputer lokal.

Setiap paragraf latar belakang diatas dapat disimpulkan, maka dari itu penulis merasa perlu dibuat aplikasi temu kembali informasi dengan menggunakan model ruang vektor.

1.2. Perumusan Masalah

Adapun yang menjadi rumusan masalah dalam Tugas Akhir ini adalah:

1. Bagaimana merancang sistem yang dapat mencari data yang dibutuhkan secara cepat dan akurat pada komputer yang stand alone?

2. Bagaimana cara kerja mesin pencari dalam melakukan pembobotan dokumen dengan menggunakan metode TF-IDF.
3. Bagaimana mesin pencarian dapat menghasilkan dokumen yang teranking berdasarkan query masukan oleh pengguna.

1.3. Batasan Masalah

Dalam penyusunan Tugas Akhir ini penulis memberikan batasan-batasan masalah sebagai berikut:

1. Data atau dokumen yang dicari hanya berformat teks (.txt).
2. Ukuran file dari data yang dicari tidak dibatasi ukurannya.
3. Pengindeksan data atau dokumen dilakukan berdasarkan perhitungan bobot.
4. Aplikasi yang dibuat menggunakan teknik pencarian Model Ruang Vektor yaitu suatu model yang digunakan untuk mengukur kemiripan antara isi suatu dokumen.
5. Pencarian kata/ keyword tidak bisa lebih dari satu kata.
6. Pencarian kata/ keyword tidak bisa mencari judul atau nama file/dokumen.
7. Aplikasi tidak bisa mencari tanda baca.

1.4 Tujuan

Tujuan dari Skripsi ini adalah untuk memudahkan pengguna untuk merancang sebuah sistem yang dapat mencari data yang dibutuhkan secara cepat pada komputer yang stand alone, maka sebab itu pengguna membuat aplikasi

perncarian kata dengan melakukan pembobotan dokumen menggunakan metode TF-IDF. Aplikasi dapat menghasilkan dokumen yang di cari setelah melakukan pencarian kata dan pengguna dapat mengetahui proses dari pembobotan dokumen dan perankingan dengan menggunakan metode TF-IDF.

1.5 Manfaat

Manfaat penelitian Tugas Akhir akhir ini adalah mempermudah pengguna untuk mencari file yang diinginkan dengan cepat.dan mempermudah pengguna mendapatkan dokumen relevan berdasarkan query yang dimasukkan, serta telah teranking berdasarkan tingkat relevansinya.

1.6 Metodolologi Penelitian

Langkah-langkah dalam pengerjaan skripsi ini antara lain:

1. Studi literatur tentang:

Konsep/metode text mining, Metode pembobotan TF-IDF dan penerapannya teori search engine.

2. Perencanaan dan pembuatan perangkat lunak modul pemrosesan dokumen modul pemrosesan query modul matching (pencocokan).
3. Pengujian dan analisa perangkat lunak Pengujian program yang telah dibuat Analisa hasil output dari program
4. Pengambilan kesimpulan Pengambilan kesimpulan dengan cara membandingkan hasil output program dengan hasil pengamatan secara manual antara query masukan dengan dokumen hasil keluaran.

1.7 Sistematika Penulisan

Sistematika dalam penyusunan Tugas Akhir ini akan dijabarkan dalam setiap bab dengan pembagian sebagai berikut :

BAB I : PENDAHULUAN

Berisi tentang latar belakang diambilnya judul Skripsi “Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) pada Sistem Temu Kembali Informasi” , rumusan masalah yang akan dikaji dalam Skripsi ini, ruang lingkup atau batasan masalah, tujuan dan manfaat dari pembuatan Skripsi ini, metodologi penelitian, dan sistematika penulisan Skripsi yang menjelaskan secara garis besar substansi yang diberikan pada masing-masing bab.

BAB II : TINJAUAN PUSTAKA

Pada bab ini dijelaskan landasan teori yang berkaitan dalam penyelesaian masalah serta teori yang mendukung dalam pembuatan sistem. Teori-teori tersebut antara lain : Pengertian Sistem Kembali Informasi, Pemodelan Rung Vektor, Pengembangan Perangkat Lunak, Pengujian perangkat Lunak, Pengertian Delphi,

BAB III : ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas tentang model penelitian dan prosedur penelitian antara lain: Analisis Sistem, Perancangan Sistem

menggunkan metode Ruang Vektor, Perancangan Database dan lain sebagainya.

BAB IV : IMPLEMENTASI PERANGKAT LUNAK

Pada bab ini akan dibahas tentang tampilan dari sistem yang telah dibuat dan dari implementasi ke bahasa pemrograman program yang telah dibuat.

BAB V : UJI COBA DAN EVALUASI PROGRAM

Dalam bab ini berisi tentang uji coba pada sistem yang telah jadi dan mengadakan evaluasi atau pengecekan terhadap sistem yang sudah berjalan.

BAB VI : PENUTUP

Pada bab ini akan dibahas tentang kesimpulan dari uji coba dan saran-saran yang diambil dari kelemahan aplikasi sebagai pengembangan lebih lanjut dari sistem yang telah dibuat.